David L. Post

Armstrong Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio 45433-6503

Christopher S. Calhoun

Systems Research Laboratories, Inc.
Dayton, Ohio 45440-3696

ASD-88-2492

An Evaluation of Methods for Producing Desired Colors on CRT Monitors

Three experiments were performed to evaluate methods for generating colors having desired CIE chromaticity coordinates and luminances on CRT monitors. Methods based on piecewise linear interpolation were found to yield better accuracy than others based on linear regression. Comparisons between two monitors revealed no substantial differences in accuracy, but the monitor's state of adjustment was found to have important effects. Evidence was also obtained concerning the number of CRT measurement points needed to obtain maximum accuracy from the interpolation-based methods. The methods' accuracies are characterized and avenues for improvement are suggested.

Introduction

An increasing number of vision researchers are using color CRT monitors, driven by digital graphics (DG) systems, to produce color stimuli. Sometimes, the colors must have specific CIE chromaticity coordinates and luminances. In such cases, these persons face a common problem: developing a method for generating specific colors on a DG/color monitor system.

If only a few colors are needed, this problem can be solved via trial-and-error. Otherwise, one must first measure the monitor's colorimetric behavior as a function of the values loaded into the DG system's digital-to-analog convertors (DACs, which output the voltages that modulate the monitor's red, green, and blue electron guns) and then use the resulting data to develop equations that predict the DAC values needed to produce specific colors. Many methods can be devised for accomplishing these two tasks, some of which have been reported[1–4] and others not. Their relative merits are unclear, however, because little effort has been made to validate the methods and quantify their predictive accuracies. The present experiments compared the accuracies of several methods for calculating the DAC values that will produce desired chromaticity coordinates and luminances on a DG/monitor system.

Seven predictive models, representing three classes of equations (polynomials, power functions, and exponentials) and two modeling approaches (regression and linear interpolation), were compared. All seven models use the assumption that the monitor's guns do not interact. This has two advantages. First, it implies that the monitor's colorimetric behavior can be adequately characterized by measuring each gun separately. Each gun is stepped through a series of DAC values while the other two guns are set to zero and the monitor is measured at each step. The required number of measurements is, thus, much less than would be the case if interactions were expected. The second advantage is that the equations do not require interaction terms, which greatly simplifies solving for DAC values.

Six of the models also assume that the guns' chromaticity coordinates are invariant. Explaining the consequences of this assumption requires discussing tristimulus colorimetry. For any color produced by additively mixing three primaries, the relationship between the primaries' luminances and the color's CIE 1931 *XYZ* tristimulus values is given by the matrix expression

$$\mathbf{T} = \mathbf{C}\mathbf{Y}, \qquad (1)$$

where $\mathbf{Y}$ is a 3×1 vector containing the primaries' luminances, $\mathbf{T}$ is a 3×1 vector containing the resulting tristimulus values, and

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **1989**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **An Evaluation of Methods for Producing Desired Colors on CRT Monitors**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory Wright Patterson AFB, OH 45433**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**15**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

$$\mathbf{C} = \begin{vmatrix} x_R/y_R & x_G/y_G & x_B/y_B \\ 1 & 1 & 1 \\ z_R/y_R & z_G/y_G & z_B/y_B \end{vmatrix}, \tag{2}$$

i.e., $\mathbf{C}$ is a 3×3 matrix where x_R is the red primary's CIE 1931 x chromaticity coordinate, etc. Equation (1) implies that

$$\mathbf{Y} = \mathbf{C}^{-1}\,\mathbf{T}, \tag{3}$$

i.e., the luminances needed to produce a desired color can be calculated by multiplying its tristimulus values by the inverse of $\mathbf{C}$. If the primaries' chromaticity coordinates are invariant, $\mathbf{C}$ will be also, which simplifies forming the matrix. Equations (1)-(3) are also valid if CIE 1976 chromaticity coordinates and tristimulus values are substituted.

The first six models described below use Eq. (3) to identify the required primary luminances. They differ only in the ways in which they then determine the DAC values that will produce these luminances. Since measurement signal-to-noise ratio usually improves as luminance increases, the chromaticity coordinates used to form $\mathbf{C}$ in the present experiments were those obtained for each gun at the largest DAC value measured during monitor calibration. The models are the following:

PLCC (Piecewise Linear interpolation assuming Constant Chromaticity coordinates): This is the easiest model to program. Each gun's luminance is assumed to change linearly below, between, and above all measured DAC values. The adequacy of this assumption depends, of course, on making the steps between measured DAC values small. Our implementation of PLCC also assumes that luminance is zero at zero DAC value.

LIN-LIN2 (linear-linear second-order model): This approach models the relationship between DAC value and luminance for each gun using the second-order polynomial $Y = a + bD + cD^2$, where D is the gun's DAC value, Y is the resulting luminance, and a, b, and c are coefficients obtained via linear regression. This model captures the curvilinear relationship between DAC value and luminance. It is more complicated than PLCC because the software must perform second-order linear regression and, when solving for DAC values, must be able to decide which solution is appropriate. (Note that both may be positive.)

LOG-LOG: The relationship between DAC voltage and luminance is often said to approximate a power function, which implies that equations of the form $\log Y = a + b \log D$ should work. This also requires a regression routine, but this routine and the DAC-value solution algorithm are simplified by the use of a first-order model. The algorithm must, however, avoid attempting to take the logarithm of zero or a negative number. Negative values for Y and/or D can arise if the desired color lies outside the monitor's color gamut.

LOG-LOG2: Cowan[12] has reported that LOG-LOG does not always yield accurate results and has recommended a second-order version of the form $\log Y = a + b \log D + c(\log D)^2$. This has the same disadvantages as LIN-LIN2 and LOG-LOG.

LOG-LIN: One can fit an exponential model using equations of the form $\log Y = a + bD$. This requires only first-order regression and has only one root, which allows easy solution for DAC values.

LOG-LIN2: One can also use a second-order version of LOG-LIN, i.e., $\log Y = a + bD + cD^2$, which has the same disadvantages as LOG-LOG2.

PLVC (Piecewise Linear interpolation assuming Variable Chromaticity coordinates): This model was introduced by Farley and Gutmann[2] and is the same as PLCC except the guns' chromaticity coordinates are assumed to change as a function of luminance. The coordinates obtained at each DAC value during monitor calibration are considered to be correct. They are assumed to change linearly at intermediate values. This model requires interpolating for all three DAC values simultaneously, which is more complicated than using PLCC. Methods for accomplishing this are discussed in Appendix A.

Method

Apparatus

The apparatus consisted of a minicomputer, digital image processor, two 19-inch (48-cm) color monitors (Tektronix 690-SR and Aydin 8830 Patriot) with 0.31-mm mask pitch, and a computer-controlled spectroradiometer. The Tektronix contains a delta-gun CRT and has achieved an especially favorable reputation among some vision researchers. The Aydin contains an inline-gun CRT and is representative of many commonly used units. The image processor contains 10-bit DACs, i.e., each DAC maps integers ranging from 0-1023 onto its output voltage range (i.e., 0-1 VDC).

Basic Procedure

For each replication of an experiment, the monitor was calibrated by measuring each gun spectroradiometrically at several equally spaced DAC values. For each gun, the first DAC value was the smallest that produced reliable measurements and the last was the largest that could be loaded into the DACs (1023). (The lowest luminance that can be reliably measured by our spectroradiometer is approximately 0.02, 0.10, and 0.04 cd/m^2 for the red, green, and blue guns, respectively. The corresponding DAC values vary according to the monitor and its adjustment.) The resulting calibration file was then used to calculate the DAC values needed to produce a standard set of colors, using various models. The standard color set consisted of 28 pairs of chromaticity coordinates (shown in Figure 1) that fully sampled the monitors' chromatic gamuts. They were replicated at 10 and 30 cd/m^2.

Each set of values was loaded into the DACs and the monitor was measured to determine the resulting chromaticity coordinates and luminance. Two dependent measures were calculated for each observation: (1) absolute value of the percent luminance error; and (2) chromatic error, expressed as distance on the CIE 1976 uniform chromaticity scale (UCS) diagram. The order of the measurements was

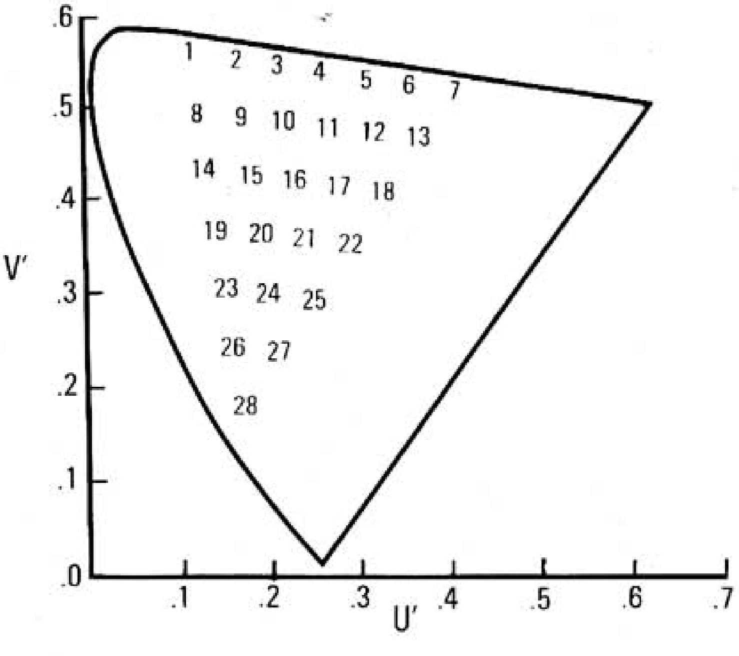

FIG. 1. CIE 1976 UCS diagram showing the 28 chromaticities used in the experiments.

randomized for each replication. Two monitors were tested, to check the generalizability of the results.

Two methods of adjusting the monitors were also tested, to evaluate the effects of monitor setup. The standard setup involved adjusting the monitor's white point to the chromaticity coordinates of CIE standard illuminant D_{65} while maximizing the blue gun's luminance.* This is fairly conventional and has the advantage of tending to maximize the DG/monitor system's colorimetric gamut. The alternate ("equiluminous") setup involved setting each gun's maximum luminance at approximately 35 cd/m^2, which is roughly the highest possible luminance from our monitors' blue guns. This sacrifices gamut but yields a consequent improvement in colorimetric resolution, because the range of possible DAC values is mapped onto a smaller luminance range for the red and green guns, i.e., the red and green DAC values are mapped more densely.

Experiment 1

The first experiment compared the seven models to identify the most accurate ones. Monitor calibration was performed as described earlier. In this case, we measured 16 DAC values per gun. Two replications were performed on each monitor.

A full-factorial analysis of variance (ANOVA) was performed for both dependent measures, using Method, Monitor, Setup, Luminance, Chromaticity, and Replication as the main effects. Statistically significant effects (significance was assessed using alpha = 0.05 in all analyses herein) account for 99.5 and 97.2% of the variance in luminance error and chromatic error, respectively. These results reflect the fact that there was little variation across the two replications, so nearly every effect is significant.

Most of the effects involving Monitor are significant, but they account for little of the variance. Overall, the Aydin yielded slightly better accuracy than the Tektronix (average absolute luminance error of 19.8 vs. 21.4% and average chromatic error of 0.014 vs. 0.015). These differences have little practical importance.

Table I shows the mean errors from the Method × Setup × Luminance interaction. It provides a satisfactory summary of the most important results, although it must be noted that effects involving Chromaticity account for much of the variation in chromatic error. Critical differences for luminance error and chromatic error in Table I were calculated using Fisher's Least Significant Difference (LSD) post hoc paired-comparison procedure. This yielded values of 0.8% and 0.001, respectively. Using these criteria, it appears that most of the differences in Table I are statistically reliable. Therefore, the models based on linear regression are clearly inferior to PLVC and can be eliminated. The same argument could be made concerning PLCC, but it is the second most accurate model, is much simpler than PLVC, and its accuracy seems to be at least

*After completing the standard setup for the Aydin, we found that the maximum luminance of the Aydin's blue gun was roughly 20 cd/m^2. This was insufficient, because our color set requires luminances as high as 30 cd/m^2 from each gun. We could not get more luminance from the red and green guns, so we increased the blue gun's gain while leaving the red and green guns alone. This means the Aydin's "white point" was actually a desaturated blue. The resulting CIE 1976 chromaticity coordinates were $u' = 0.197$, $v' = 0.401$, which corresponds to a correlated color temperature of 34,259 K.

TABLE I. Mean errors from Experiment 1.

Model	Standard setup				Equiluminous setup			
	10 cd/m²		30 cd/m²		10 cd/m²		30 cd/m²	
	$\|\%Y\|$	$u'\ v'$	$\|\%Y\|$	$u'\ v'$	$\|\%Y\|$	$u'\ v'$	$\|\%Y\|$	$u'\ v'$
PLCC	1.9	0.008	2.2	0.005	5.6	0.011	4.7	0.007
PLVC	2.0	0.004	2.2	0.003	4.5	0.006	4.2	0.004
LIN-LIN2	10.1	0.017	3.7	0.008	9.7	0.013	4.1	0.006
LOG-LOG	46.8	0.015	22.3	0.017	49.5	0.013	10.6	0.027
LOG-LOG2	8.2	0.017	7.0	0.010	8.9	0.022	11.4	0.012
LOG-LIN	108.5	0.020	69.3	0.026	71.0	0.015	18.2	0.034
LOG-LIN2	44.6	0.024	13.6	0.028	16.4	0.025	15.5	0.023

marginally acceptable for some applications. Therefore, PLCC and PLVC were retained for further study.

Table II shows the average R^2 values obtained from the regression-based models for each monitor setup during the experiment. It can be seen that their rank order corresponds fairly well with the models' relative performance. However, the models' accuracies are lower than might be expected, given the R^2s Farley and Gutmann[2] and Neri[4] have also noted that R^2s which would ordinarily be regarded as very good do not necessarily assure high accuracy.

Better understanding of the regression-based models' performance can be obtained by plotting their regression lines. Figure 2 shows an example, derived from an Aydin blue gun calibration. Initially, one notices that LIN-LIN2 provides a much better fit than the others. In fact, everything else is rather poor, despite the R^2s. This is because the others were fit (and the R^2s obtained) in transformed domains. Of particular interest is the fact that the error which was minimized in these cases was the log luminance error. Considering the dependent measures used in Experiment 1, one might expect this to yield better results than minimizing luminance error. However, given LIN-LIN2's accuracy relative to the other models', this expectation is apparently erroneous.

It is possible to fit the other models without transforming the data. This, however, requires nonlinear regression methods, which means one must either obtain the coefficients separately or add significant complexity to the DAC-value solution routine. To investigate the promise of this approach, the Statistical Analysis System's[5] NLIN procedure (multivariate secant method) was used to compute coefficients for the data shown in Fig. 2, using the nontransformed equivalents of the LOG-LOG, LOG-LOG2, LOG-LIN, and LOG-LIN2 models.* The resulting R^2 for LOG-LIN is 0.998, but the visual fit is not very good. The other models yield R^2s better than 0.999 and provide good visual fit. Therefore, it seems justifiable to explore them further, using this approach. Whether this will yield better accuracy than PLCC (or even LIN-LIN2) is an open question, but Neri's[6] report that even sixth-order polynomials are inferior to PLCC suggests otherwise.

TABLE II. Mean R^2 values from regression models.

Model	Setup	
	Standard	Equiluminous
LIN-LIN2	0.999	0.999
LOG-LOG	0.957	0.909
LOG-LOG2	0.993	0.982
LOG-LIN	0.800	0.834
LOG-LIN2	0.940	0.957

For PLCC and PLVC, the standard setup yielded better accuracy than the equiluminous setup. Therefore, the remainder of this article only reports results obtained using the standard setup. (Earlier experiments using the equiluminous setup have also yielded results for PLCC and PLVC which are inferior to the standard setup results reported here.[7])

An attempt was made to ascertain the reason for the standard setup's superiority. The calibration files were used to create plots of luminance as a function of DAC value for each gun, but this revealed no obvious irregularities for the equiluminous setup. That is, there is no evidence that reducing the red and green maximum luminances (i.e., reducing gain) caused the guns to behave peculiarly.

The only noteworthy difference is that the red and green luminances rise much more rapidly for the standard setup, which means the smallest measurable red and green DAC values are smaller. For example, for the Tektronix with the equiluminous setup, the smallest measurable red and green DAC values were 258 and 303, respectively, whereas they were both 123 with the standard setup. Since the maximum red and green luminances were much higher with the standard setup, though, it appears that, contrary to expectation, the density of the mapping of DAC value onto luminance was not important. Instead, the important factor may be that the standard setup lead to less extrapolation below the measured range when solving for DAC values. This does not mean that mapping density is never of consequence. It may simply be that our 10-bit DACs provided more than was needed.

*These are, respectively: $Y = aD^b$, $Y = aD^{(b+c \log D)}$, $Y = 10^{(a+bD)}$, and $Y = 10^{(a+bD+cD^2)}$.

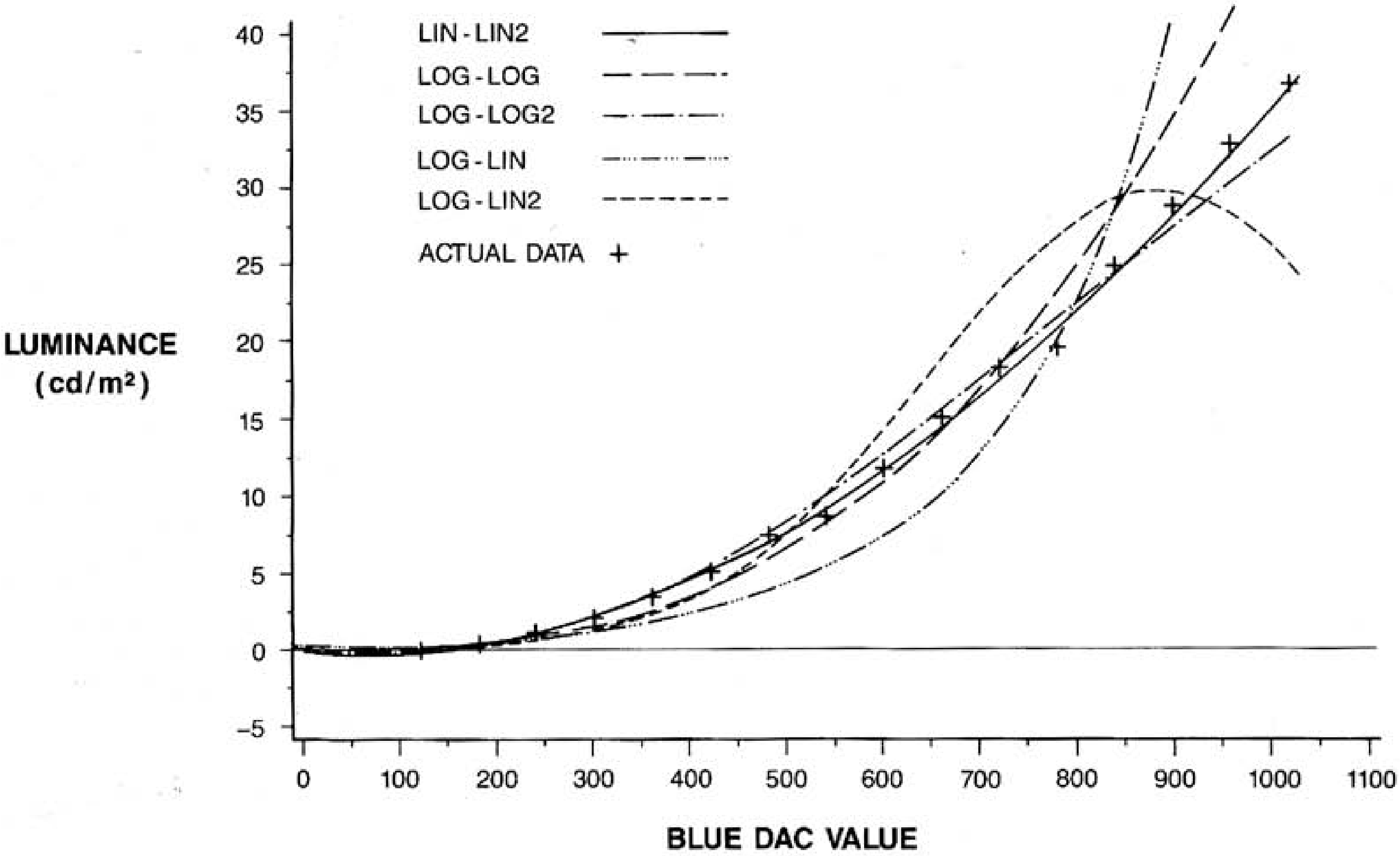

FIG. 2. Luminance as a function of DAC value for Aydin blue gun, as measured (individual crosses) and as predicted by each linear regression model (solid and broken lines).

Experiment 2

The next step was to perform a more thorough evaluation of PLCC and PLVC. This was accomplished by expanding the standard color set to include a replication at 3 cd/m² and by testing the effect (which will be termed "resolution") of the number of measurements contained in the monitor calibration file. The latter test involved: (1) measuring 32 equally spaced DAC values per gun during calibration, thereby obtaining a 32-point calibration file; (2) deleting every other measurement to obtain a 16-point file; (3) repeating to obtain an 8-point file; and (4) using PLCC and PLVC to calculate DAC values, using each file. Two replications of the experiment were performed on each monitor.

ANOVAs were performed as before, but with the main effect of Resolution substituted for Setup. Once again, nearly every effect is statistically significant, accounting for a total of 98.9 and 96.9% of the variance in luminance error and chromatic error, respectively.

The most important results are summarized in Table III, which shows the Luminance × Resolution × Method interaction. The LSD critical differences for luminance error and chromatic error in this table are 0.8% and 0.001, respectively.

There are several points worth making here. First, 16-point resolution is clearly superior to 8-point and seems to provide as much accuracy as can be obtained. Indeed, it is even superior to 32-point resolution, at 3 cd/m². The effect of resolution seems to diminish as luminance increases, though—perhaps because computations for the more luminous colors tend to involve the upper range of the DAC value versus luminance functions, where there is less curvature.

Second, accuracy improves as luminance increases. Luminance error seems to stabilize somewhere between 3 and 10 cd/m². PLVC's chromatic error behaves similarly, but PLCC's shows substantial improvement between 10 and 30 cd/m².

Third, PLCC and PLVC yield comparable luminance er-

TABLE III. Mean errors from Experiment 2.

		PLCC		PLVC	
Lum	Res	$\|\%Y\|$	$u'\ v'$	$\|\%Y\|$	$u'\ v'$
3	8	20.3	0.042	19.8	0.044
	16	4.2	0.015	4.9	0.008
	32	6.2	0.018	7.3	0.009
10	8	5.4	0.010	5.1	0.007
	16	2.0	0.009	1.9	0.004
	32	2.3	0.009	2.3	0.004
30	8	2.7	0.006	2.7	0.004
	16	2.4	0.005	2.2	0.003
	32	2.2	0.005	2.2	0.003

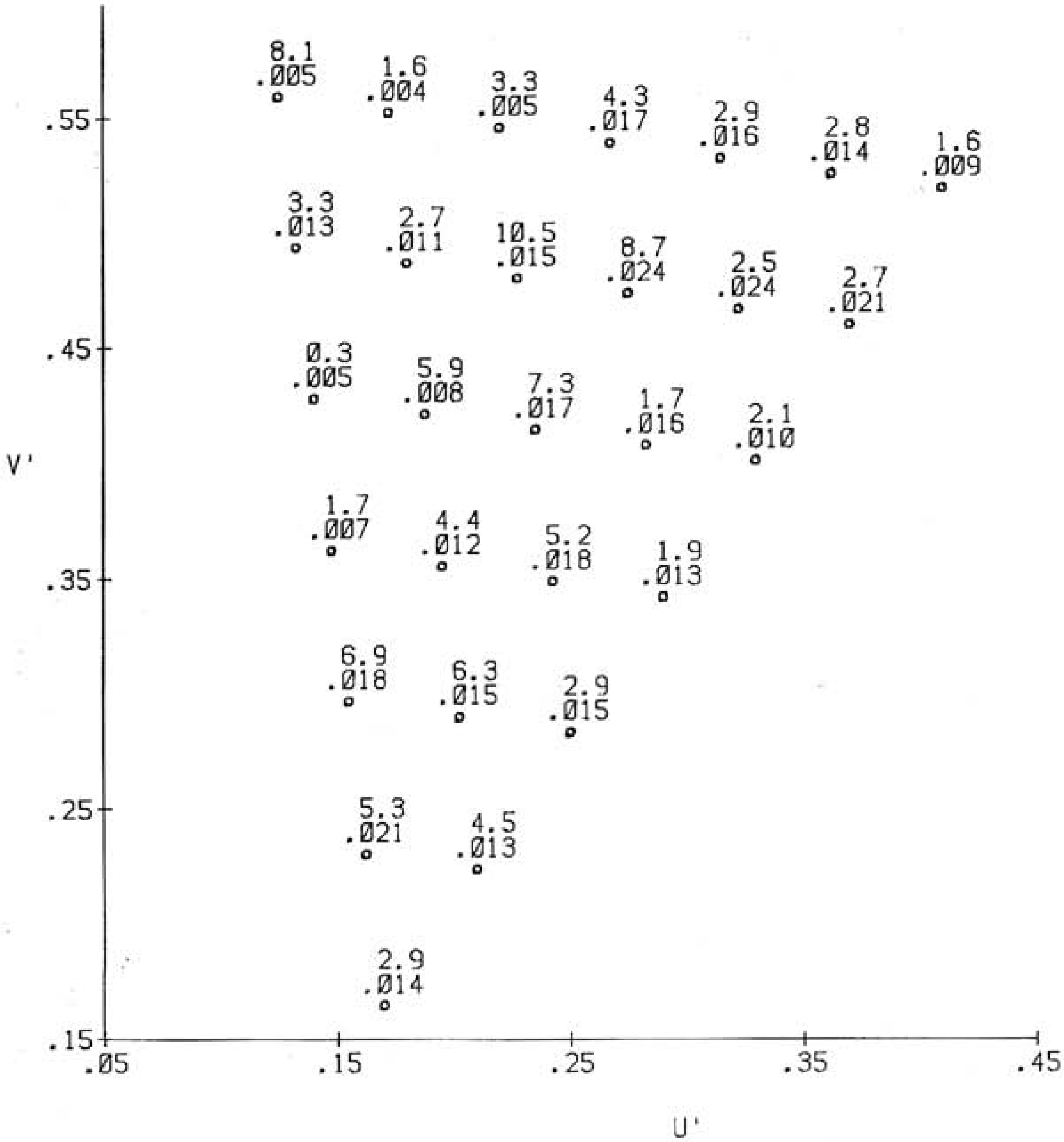

FIG. 3. CIE 1976 UCS diagram inset showing mean absolute percent luminance error (upper values) and mean chromatic error (lower values) from Experiment 2 for Aydin monitor using 16-point calibration resolution. PLCC at 3 cd/m².

ror, but PLVC clearly yields less chromatic error, although the differences diminish as luminance increases. These findings are logical because (1) PLCC and PLVC model gun luminance in the same way, (2) shifts in the guns' measured chromaticity coordinates occur mainly at low luminances, and (3) PLCC's estimates of gun chromaticity are derived from the maximum luminances. Therefore, PLVC's DAC value solutions approach PLCC's, as luminance increases.

Finally, it is worth pointing out that the values at 10 and 30 cd/m^2 for 16-point resolution are very similar to the comparable values in Table I. This provides added confidence in the repeatability of our results.

The effects of Chromaticity are illustrated in Figs. 3–8. The figures show chromatic and luminance error for PLCC and PLVC at 3, 10, and 30 cd/m^2, averaged across the two replications, for the Aydin using 16-point resolution. They are intended to illustrate the range of accuracies one might expect to achieve. They also underscore the fact that, although either model can yield good mean accuracy plus excellent accuracy for specific colors, they can also yield substantial error in some cases. The colors for which error is high are not always the same on the two monitors, but there are many similarities.

Effects involving Monitor account for greater proportions of the variance than in Experiment 1. For PLVC, Monitor makes little overall difference. For PLCC, however, chromatic accuracy tends to be better on the Aydin (average of 0.011 vs. 0.015), although the differences diminish as luminance increases. These trends are also evident in the standard-setup results from Experiment 1, but not in the equiluminous-setup results.

We believe the effects of Monitor are due mainly to changes in the Tektronix' chromatic gamut which occurred during the course of the experiments. Its CRT was replaced shortly before we started data collection, and the equiluminous-setup condition was run first. Our chromaticity set

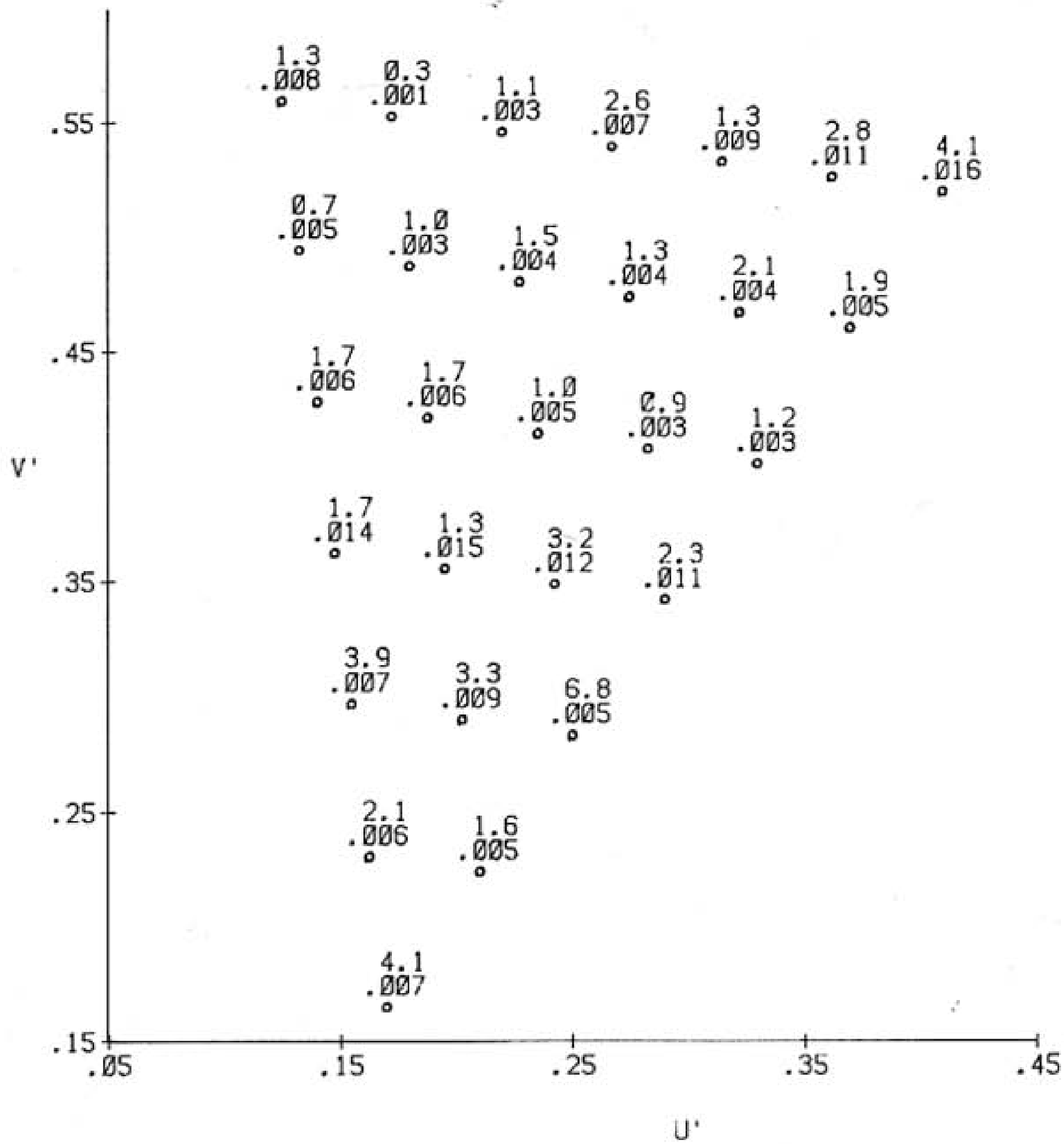

FIG. 4. Same as Fig. 3, PLCC at 10 cd/m².

was originally designed to fall within both monitors' gamuts at all luminances,* but apparently deviated from this criterion as the Tektronix CRT burned in. The resulting problems, which manifest themselves most obviously in the form of increased chromatic error at the gamut edges, are evident for both PLCC and PLVC. PLVC, however, tended to yield lower chromatic error elsewhere for the Tektronix and, therefore, its average error was comparable across Monitor.

*Although a monitor's chromatic gamut is typically portrayed as a simple, invariant triangle on the chromaticity diagram, it is actually much more complex. One reason is that the guns have luminance limitations, so the gamut changes shape and collapses as these maxima are exceeded. Furthermore, the guns' measured chromaticity coordinates vary with luminance. Farley[8] and Post[9] have suggested that this is due primarily to signal-to-noise ratio problems in the measuring device. It is also likely, though, that phosphor persistence, unintentional excitation by room lights and reflections off the faceplate, imperfect purity adjustment, shadowmask motion, and stray electrons play a role.

Experiment 3

A third experiment was performed to test PLCC and PLVC at 100 cd/m². Because of limitations on each gun's maximum luminance, it was not possible to use the complete set of 28 chromaticities in this test. Therefore, a subset consisting of chromaticities 1, 2, 3, 8, 9, 10, 14, 15, 16, 19, and 20 (see Fig. 1) was used. Two replications of the experiment were performed on each monitor using 16-point calibration resolution.

TABLE IV. Mean errors from Experiment 3.

	PLCC			PLVC		
Lum	%Y	|%Y|	u' v'	%Y	|%Y|	u' v'
3	1.2	3.8	0.011	0.8	4.2	0.006
10	0.0	1.1	0.008	−0.5	1.1	0.002
30	1.2	1.9	0.006	0.6	1.7	0.002
100	−0.9	2.3	0.003	−1.1	2.1	0.002

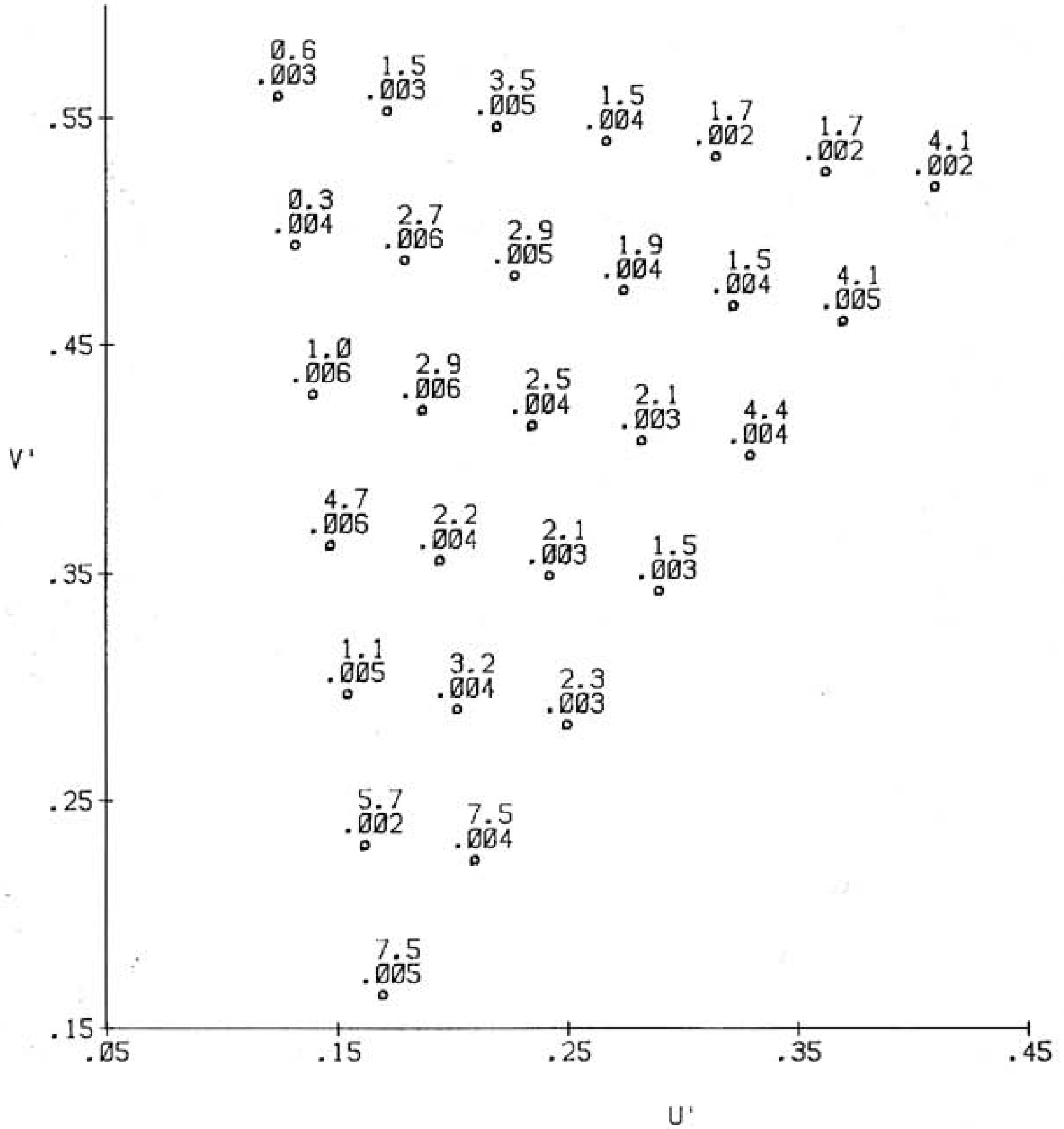

FIG. 5. Same as Fig. 3, PLCC at 30 cd/m^2.

The results are summarized in Table IV. For comparative purposes, means for the same subset of chromaticities at 3, 10, and 30 cd/m^2 are also shown. They were computed using the data from Experiment 2. In addition, the column showing absolute percent luminance error has been supplemented with another, showing percent luminance error. The LSD critical differences for luminance error, absolute luminance error, and chromatic error in Table IV are 1.3%, 1.0%, and 0.001, respectively.

The results are mostly in keeping with expectations based on the preceding findings. PLVC's chromatic error stabilizes between 3 and 10 cd/m^2 and remains stable up to 100 cd/m^2. PLCC's chromatic error continues improving up to 100 cd/m^2, at which point it is no longer reliably worse than PLVC's. Absolute luminance error does not differ reliably across models at any luminance. However, in both cases, it is reliably worse at 100 cd/m^2 than at 10 cd/m^2, and is intermediate at 30 cd/m^2. Apparently, it reaches a minimum somewhere near 10 cd/m^2 and then slowly increases. Luminance error shows an interesting and statistically significant tendency to go from being positive at 30 cd/m^2 to negative at 100 cd/m^2. This suggests subadditivity among the guns, possibly caused by limitations in the monitors' power supplies.

Most of the average errors shown in Table IV for PLVC are well within customary needs. Indeed, we wondered whether substantially better results are possible, given the inherent variability of the DG/monitor/spectroradiometer systems. Therefore, we assessed this variability.

PLVC was used to calculate Aydin DAC values for the chromaticity subset at all four luminances. The DAC values were loaded and measured five times in random order and the average luminance and chromaticity coordinates obtained for each set of DAC values was calculated. We then computed the difference between each measurement and the associated average values. Specifically, we calculated percent luminance error and error in CIE u' and v'. Next, the standard deviation of these errors was computed for each

TABLE V. Average standard deviations from Experiment 3.

	Aydin			Tektronix		
Lum	%Y	u'	v'	%Y	u'	v'
3	1.4	0.0006	0.0006	0.5	0.0003	0.0004
10	0.6	0.0003	0.0003	0.5	0.0004	0.0005
30	0.7	0.0003	0.0003	0.9	0.0002	0.0003
100	2.0	0.0003	0.0003	1.9	0.0002	0.0002

set of DAC values. (The average errors are, of course, zero.) Finally, the standard deviations were averaged to yield a summary statistic for each luminance level. This procedure was then repeated for the Tektronix.

The resulting averages are shown in Table V. Evidently, system variability constitutes an increasing proportion of the absolute luminance errors shown in Table IV, as luminance increases. Significant luminance-error reduction appears possible at 3 cd/m^2, but the potential is more modest at 10 and 30 cd/m^2 and perhaps negligible at 100 cd/m^2. PLVC's chromatic error is roughly ten times the standard deviations, which are comparable with the smallest steps in chromaticity we can take, given our 10-bit DACs. This indicates that substantial reductions in chromatic error can probably be achieved at all luminances.

There are two other points worth making about Table V. First, chromatic stability tends to improve as luminance increases. This may partially explain the reduction in chromatic error that we observed as luminance increased. Second, with the possible exception of luminance stability at 3 cd/m^2, there are no pronounced differences between monitors.

Discussion

Our findings have several practical implications for persons who wish to use DG/monitor systems to generate colorimetrically calibrated stimuli. First, PLCC and PLVC are clearly better than the other models considered here because they are more accurate. Choosing between them requires

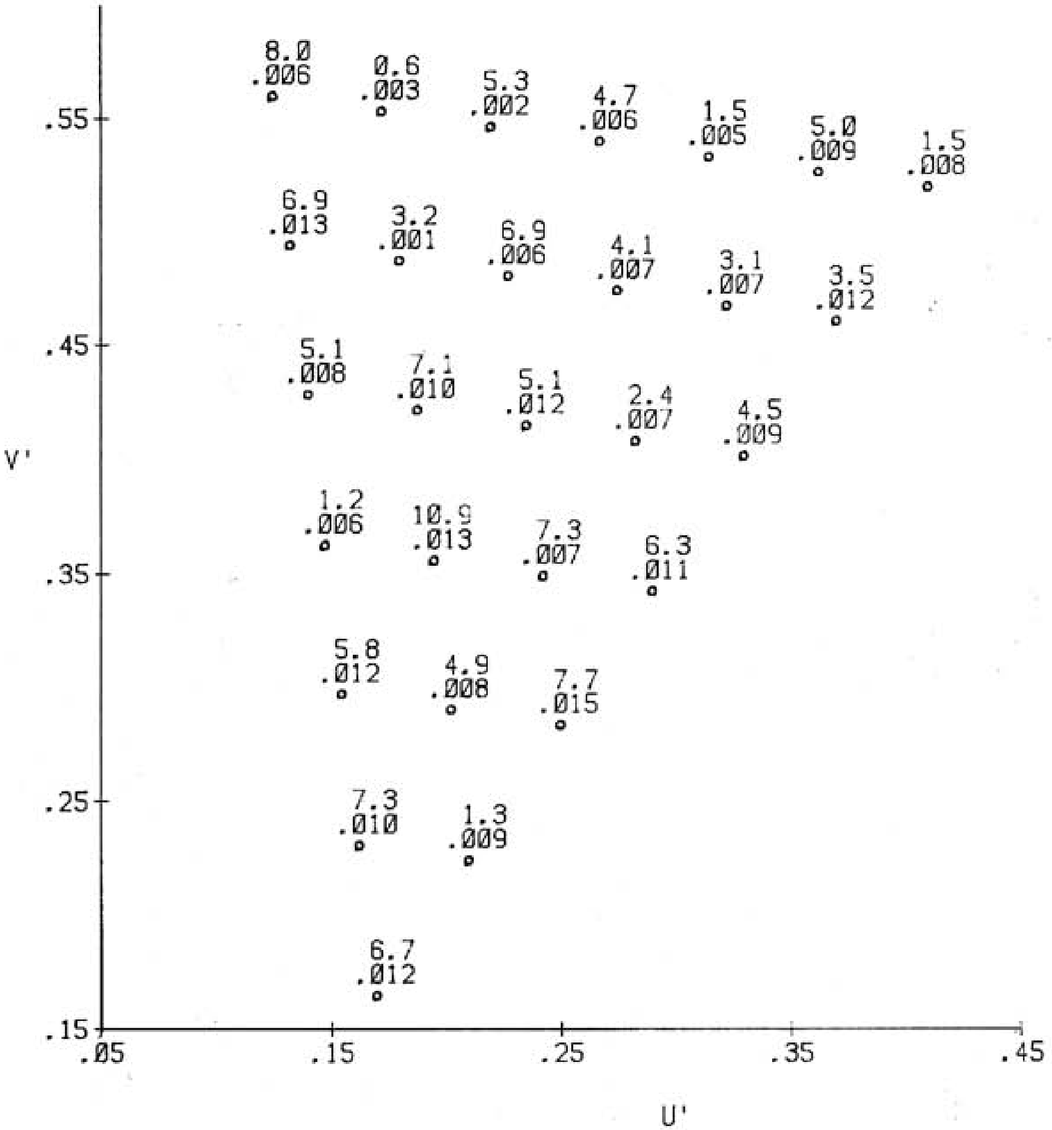

FIG. 6. Same as Fig. 3, PLVC at 3 cd/m^2.

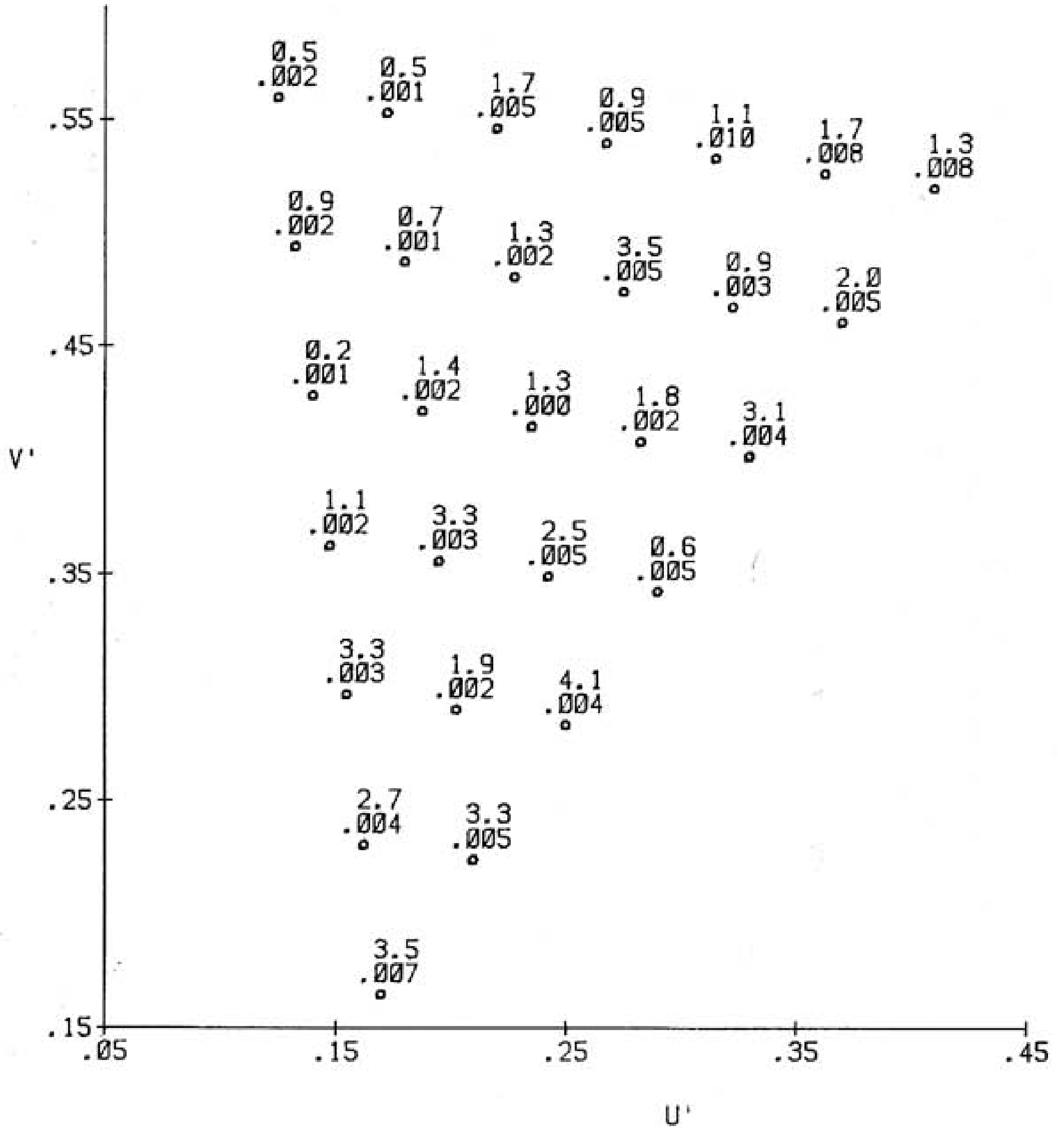

FIG. 7. Same as Fig. 3, PLVC at 10 cd/m².

considering the application. If approximations will suffice or only high luminances are needed, PLCC is better because it is easy to implement and fairly accurate. If maximum predictive accuracy across a broad range of luminances is required, though, PLVC is definitely better.

Recently, Cowan and Rowell[10] and Brainard[11] have presented analyses from which they conclude that their CRT monitors meet (within reasonable limits) the assumption that the chromaticity coordinates associated with each gun are invariant. They have used the term "phosphor constancy" to refer to this assumption (although there is no special reason to doubt that the *phosphors'* chromaticity coordinates are constant). Clearly, our conclusions regarding the merits of PLCC versus PLVC imply that our monitors behave differently. This apparent contradiction may, however, reflect differences in analyses, rather than differences in hardware. We have actually quantified the improvement produced by allowing for variable chromaticity coordinates and found it to be nontrivial. The aforementioned authors might have reached the same conclusion, if they had performed similar analyses.

Our second major finding is that, for either model, a conventional monitor setup can be expected to yield better accuracy than the alternative explored here. Possibly, even better results can be obtained by deviating from our standard setup. In particular, maximizing each gun's luminance seems worth exploring because this should reduce extrapolation below the measured range when solving for DAC values.

Third, for either model, a 16-point calibration should yield as much overall accuracy as can be achieved. Because the upper portions of the DAC-value versus luminance functions are nearly linear, though, it is likely that fewer measurements will suffice if the measurement points are chosen properly. (It must be noted, however, that monitors exhibit deceleration at the top end if the gain is set too high.) Therefore, if the monitor setup is relatively permanent, a reasonable procedure might be to initially perform a 16-point calibration, examine the results, and reduce the num-

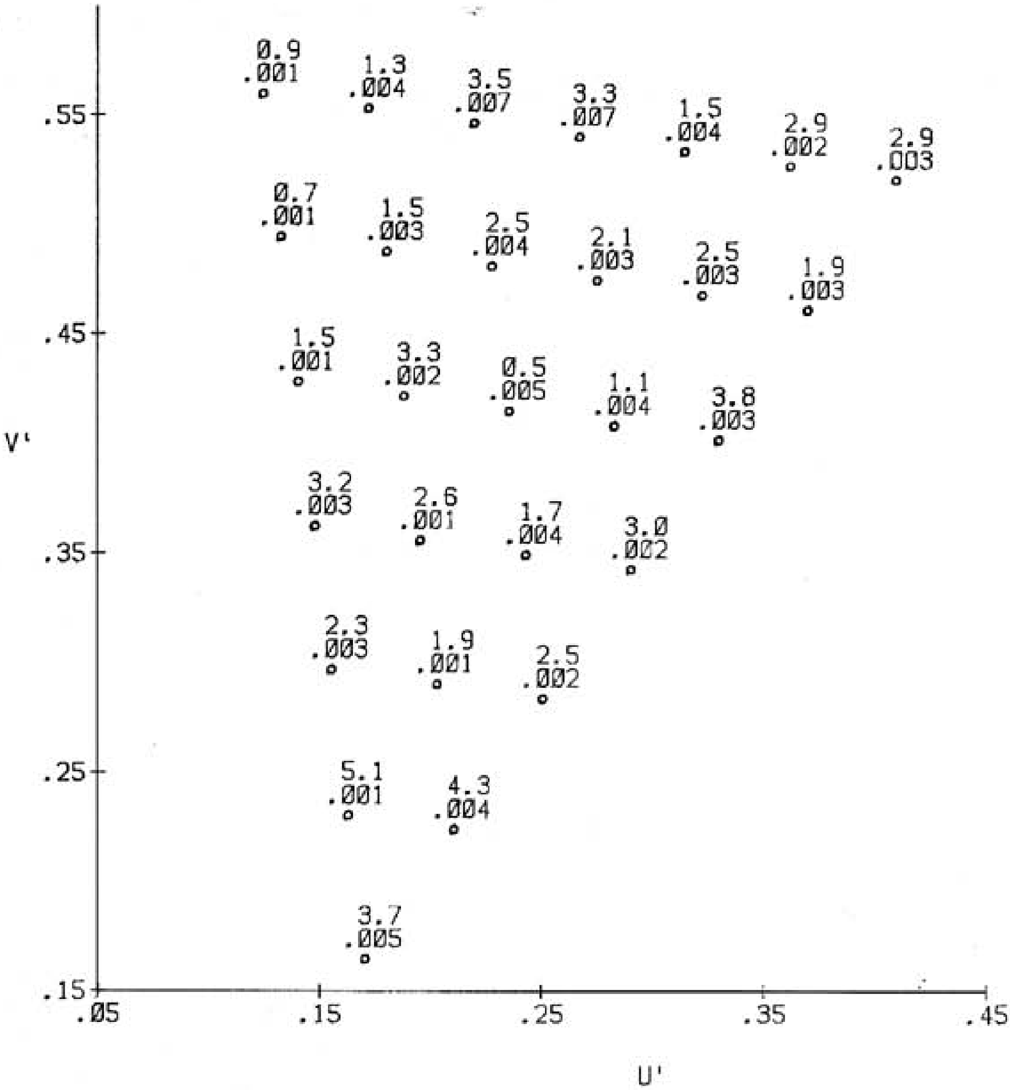

FIG. 8. Same as Fig. 3, PLVC at 30 cd/m².

ber of measurements at the upper portions accordingly in subsequent calibrations. Whether the resulting time savings will be worth this trouble depends on the application.

Fourth, neither model is very accurate at luminances of 3 cd/m² or (presumably) less. This may be inherent, but seems more probably attributable to our ability to accurately calibrate the monitors at low luminances. At worst, though, given a standard setup and 16-point calibration, PLVC's accuracy seems to stabilize somewhere between 3 and 10 cd/m² and holds reasonably well to at least 100 cd/m². PLCC's overall accuracy seems to improve with luminance up to roughly 100 cd/m², at which point it becomes equivalent with PLVC.

Finally, the fact that two very different monitors yielded very similar results suggests that our conclusions apply to many, if not most, other monitors. Indeed, it appears likely that similar accuracies would be obtained in many cases, given comparable measuring conditions.

Conclusions

The accuracies obtained here are acceptable for many applications, but not all. Substantial improvements are apparently possible (especially at the lower luminances) and greater consistency is desirable. Our evidence suggests that better characterization of the low end of the DAC value versus luminance functions would yield worthwhile improvements.

Further error reduction may be possible using models that account for interactions among the monitor's guns. Unfortunately, this requires much more complex DAC value solution algorithms and calibration procedures. In any case, though, it appears that alternate models must account for shifts in the guns' measured chromaticity coordinates if they are to outperform PLVC. It is hoped the results presented here will serve as useful benchmarks for the development of improved models.

It is important to realize that, thus far, no model has been shown to consistently yield high predictive accuracy. There-

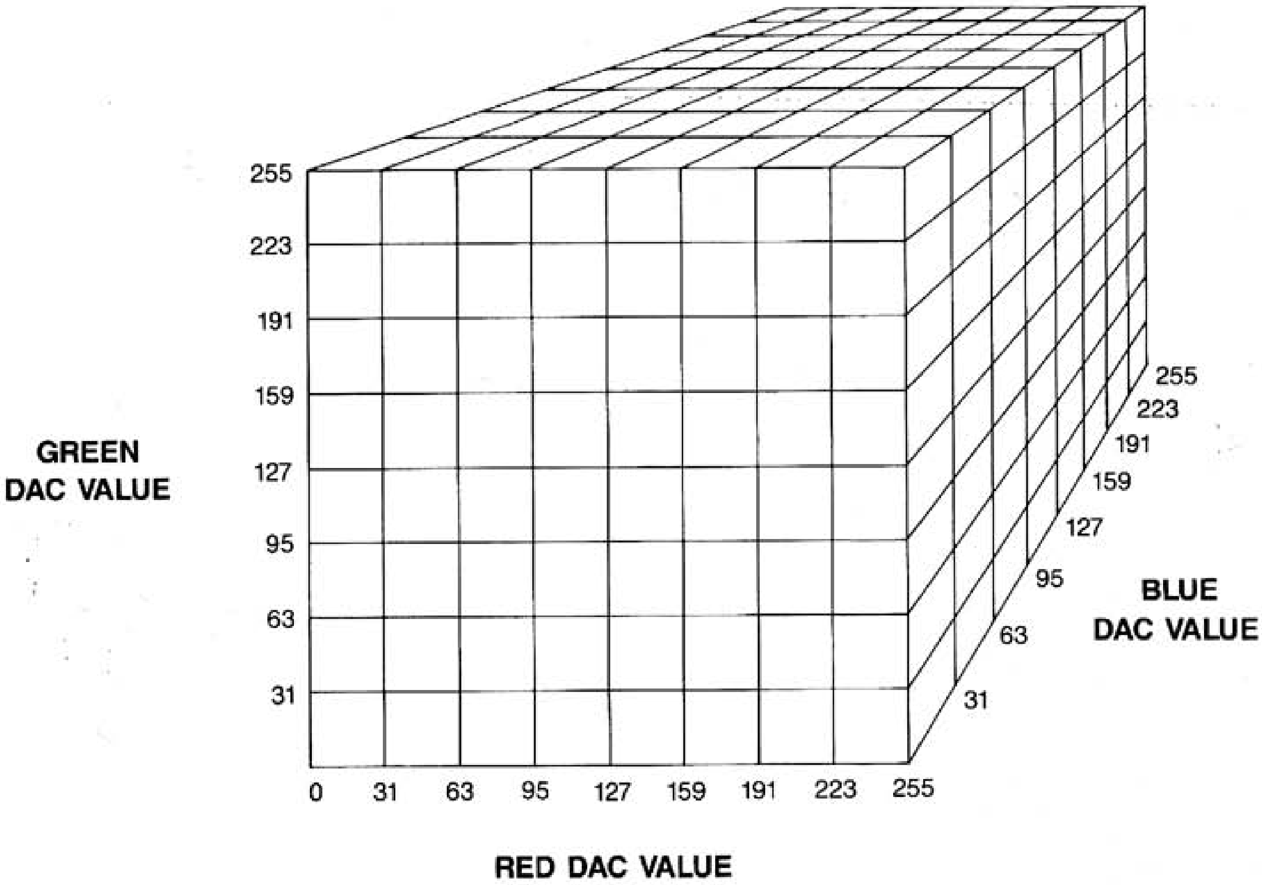

FIG. 9. PLVC solution spaces for an 8-point calibration on a system having 8-bit DACs.

fore, for applications requiring maximum accuracy, DAC value calculations should not be trusted. Instead, they should be measured and corrected. This can be accomplished by a computerized search program that uses the calculations as a starting point and then repetitively measures the resulting colors using a computer-controlled spectroradiometer and adjusts the DAC values until user-specified colorimetric criteria are met. An algorithm for accomplishing this is presented in Appendix B.

Appendix A: PLVC

Before discussing PLVC, it is worthwhile to consider alternate approaches that also allow for variable chromaticity coordinates. One is to develop regression equations, relating the elements of $\mathbf{C}^{-1}$ to $\mathbf{T}$. Given $\mathbf{T}$ and these equations, $\mathbf{C}^{-1}$ can be computed, Eq. (3) can be solved for $\mathbf{Y}$, and the DAC values needed to produce $\mathbf{Y}$ can be obtained using any of the first six methods described in the Introduction, as well as others. The main problem with this approach is that the guns' measured chromaticity coordinates do not change linearly with luminance. Instead, they change most rapidly at low luminance and then stabilize. This means the relationship between $\mathbf{C}^{-1}$ and $\mathbf{T}$ is nonlinear in a way that can be difficult to model accurately in a regression equation. Furthermore, the calculations needed to obtain the tables of independent and dependent variables for regression are numerous and cumbersome.

A simpler approach is to fit regression equations to the calibration data, independently predicting each tristimulus value for each gun as a function of DAC value. The equations can be used to create a lookup table, showing predicted tristimulus values for every possible DAC value for each gun. This table, in turn, can be used to create another table, showing predicted tristimulus values for every possible combination of DAC values. Solving for DAC values, then, becomes a relatively trivial matter of searching the second table for the closest tristimulus values and reading off the associated DAC values.

One problem with this approach is the size of the tables. If eight-bit DACs are used, the first table will contain ($3 \times 3 \times 2^8 =$) 2304 values and the second will contain ($3 \times 256^3 =$) 50,331,648. For ten-bit DACs, the sizes become 9216 and 3,221,225,472, respectively. Thus, the second table is too large to reside in memory and can be too large for a single disk, even if one is willing to suffer the time penalty for performing the disk reads. Of course, the need for the second (and even the first) table can be eliminated by doing the calculations "on the fly" but, in any case, the time required to obtain a solution on most computers will be measured in minutes, if not hours, and therefore becomes excessive for many applications. A second

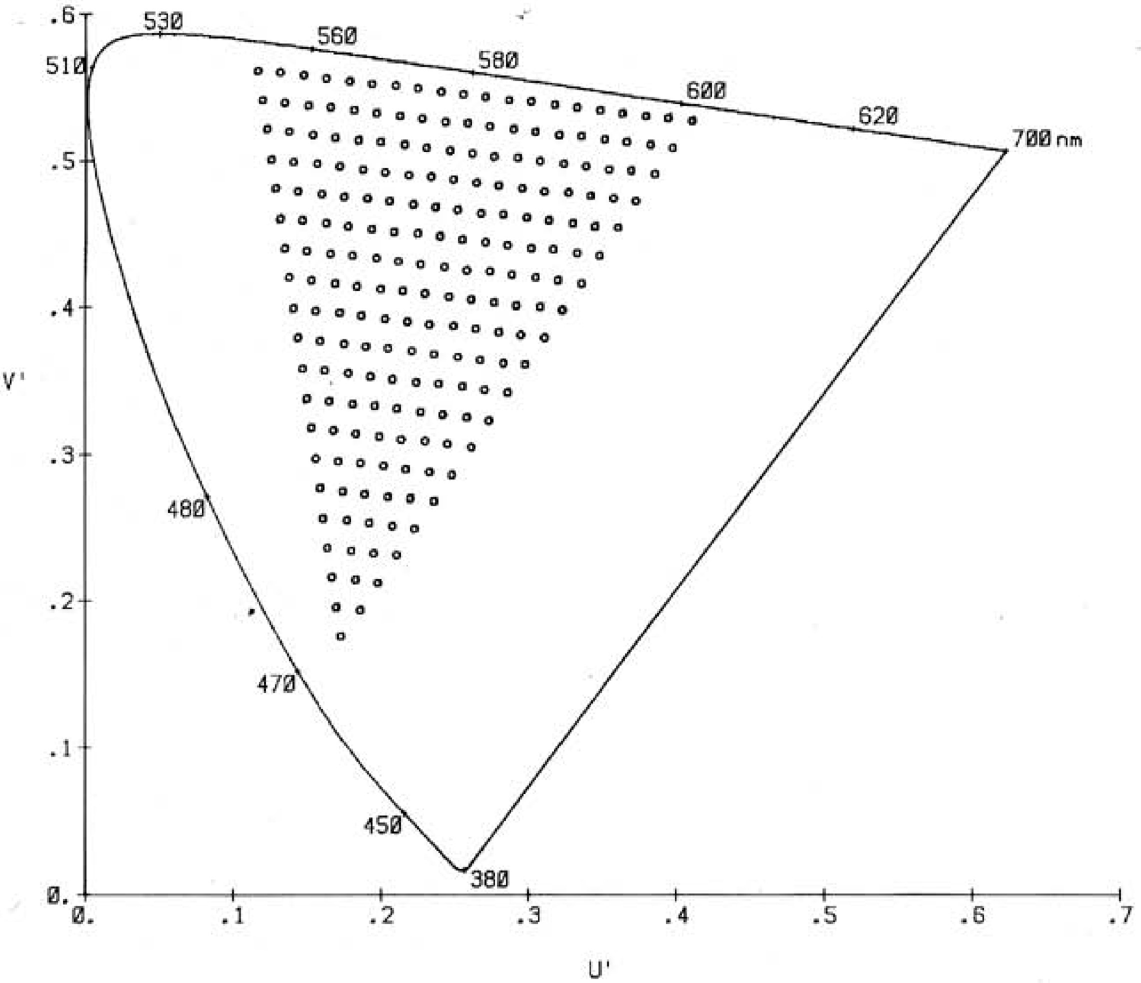

FIG. 10. CIE 1976 UCS diagram showing the 210 chromaticities adjusted each night.

problem is that our experimental results indicate that models based on linear regression are not very accurate, although the incorporation of variable chromaticity coordinates might help.

PLVC makes an easier approach possible. To explain this approach, let us simplify matters initially by assuming that the luminance of each gun is proportional with its DAC value. Note that the elements of **C** are the slopes of the guns' tristimulus values as a function of Y. We can now write a modified version of Eq. (1):

$$\mathbf{T} = \mathbf{S}\mathbf{D}, \tag{4}$$

where **D** is a 3 × 1 vector containing the DAC values, **T** is the same as before, and

$$\mathbf{S} = \begin{vmatrix} dX_R/dD_R & dX_G/dD_G & dX_B/dD_B \\ dY_R/dD_R & dY_G/dD_G & dY_B/dD_B \\ dZ_R/dD_R & dZ_G/dD_G & dZ_B/dD_B \end{vmatrix}, \tag{5}$$

i.e., **S** is a 3 × 3 matrix where dX_R/dD_R is the slope of the X tristimulus value as a function of red DAC value, etc. Thus, **S** is equal to **C**, multiplied by the constants of proportionality relating **Y** and **D**. This leads to

$$\mathbf{D} = \mathbf{S}^{-1}\,\mathbf{T}, \tag{6}$$

which is a modified version of Eq. (3). The most important difference is that Eq. (6) yields DAC values directly.

The problem in actual practice, of course, is that **Y** and **D** are not proportional and, furthermore, the guns' chromaticity coordinates vary with **Y**. Within the context of PLVC, however, Eqs. (4) and (6) can be treated as being valid within each range of interpolation. That is, we assume that, within each interpolation range, **Y** and **D** are linearly related (this assumption is also used for PLCC) and the guns' chromaticity coordinates change linearly with **D**, although the slopes vary from one range to another. Thus, we divide the DG/monitor system's colorimetric gamut into many small solution spaces, each having its associated matrix **S**, and pretend the system is linear within each space. Figure 9 illustrates this subdivision for an 8-point calibration on a system having 8-bit DACs.

If the correct space is known, the slopes that make up **S** can be readily computed from the calibration file. The only remaining complication is that intercepts must be subtracted from **T** before solving Eq. 6 and intercepts must be added to **D** afterwards (just as is the case when using PLCC). This

is fairly easy, though, because these intercepts are entries in the calibration file, anyway. Thus, in actual practice, Eqs. 4 and 6 become

$$\mathbf{T} = \mathbf{S}(\mathbf{D}\text{-}\mathbf{D}_0) + \mathbf{T}_0 \text{ and} \tag{7}$$

$$\mathbf{D} = \mathbf{S}^{-1} (\mathbf{T}\text{-}\mathbf{T}_0) + \mathbf{D}_0, \tag{8}$$

respectively, where $\mathbf{D}_0$ is a 3×1 vector containing the solution space's intercepts (i.e., the smallest DAC values within the solution space), $\mathbf{T}_0$ is a 3×1 vector containing the tristimulus values associated with $\mathbf{D}_0$, and the other terms are as before. Note that the slopes of all three tristimulus values can change independently from one $\mathbf{S}$ to another, thereby accounting for changes in each gun's measured chromaticity coordinates. (If the coordinates were constant, the slopes of X and Z would always be in the same proportions with the slopes of Y, although the slopes of Y could change independently from one $\mathbf{S}$ to another.)

It is important to realize that every solution space will yield a solution to Eq. (8). How, then, can the correct space be recognized? The correct space is the one that yields values for $\mathbf{D}$ that lie within the space. If one or more values lie outside, they have been obtained via extrapolation, rather than interpolation.

The more difficult problem is finding the correct space. Because interpolation is always performed between the two nearest measured points, the total number of solution spaces is equal to the cube of the calibration resolution. Therefore, a 16-point calibration file, for example, provides 4096 possible solution spaces. Thus far, we have been unable to devise an efficient way to directly identify the correct space, given $\mathbf{T}$. We believe it is possible to generate a lookup table, but this does not appear easy and, for a 16-point calibration file, the table requires ($3 \times 4 \times 4096 =$) 49,152 values (each entry consisting of one tristimulus value plus three DAC values). Thus, this approach does not seem very attractive. We welcome suggestions from anyone having further insight concerning this problem.

Two basic search techniques are possible. The simplest is brute force: calculate all possible solutions until the correct space is encountered. This is slow and inefficient. A better technique is to use the results from each trial solution as an indication of which space should be tried next. This usually locates the correct space very quickly, but is vulnerable to infinite loops. We use the latter technique, but have had to include code to trap and recover from loops.

It is worthwhile to point out a pitfall that can arise with either search technique. If the relationship between DAC value and the tristimulus values is not perfectly monotonic in the calibration file, multiple solutions become possible for some tristimulus values. That is, for some tristimulus values, there will be more than one solution space that yields values for $\mathbf{D}$ that lie inside the space. The most obvious way in which this can occur is if the monitor's luminance drifts too much (given the calibration resolution) during calibration. This can yield a measured luminance at one DAC value that is higher than at the next larger measured DAC value, thereby yielding a negative slope for Y (and, probably, X and Z) within that interpolation range. A less obvious mechanism involves shifts in the measured chromaticity coordinates. This can cause the slope of X and/or Z to become negative, even though the slope of Y may be positive. Therefore, if one chooses to check the calibration file for non-monotonicity, all three tristimulus values must be checked—checking Y alone is not sufficient.

One final pitfall is worth discussing. Occasionally, it will be found that no solution space meets the criterion for correctness, even though the color lies within the system's gamut. This occurs sometimes for colors that lie along the edges separating the solution spaces and is due to rounding and/or truncation in the calculations. The basic problem, then, is that one solution space indicates that the color lies just within an adjacent space, but that adjacent space indicates that the color lies just within the other. This is a special case of the infinite loop problem which was mentioned earlier and is easy to trap. Choosing between the solutions is rather arbitrary, because the associated DAC values rarely differ by more than 1. A reasonable selection rule is to choose the solution yielding the smaller predicted colorimetric error.

Appendix B: A "Measure and Adjust" Algorithm for Producing Desired Colors on Displays

We have devised a simple algorithm for the computerized search mentioned at the end of our article. Our experience has been that it is efficient and robust:

1. Calculate the DAC values needed to produce the desired color. Call these the "calculated" values.
2. Load the calculated values, measure the display, and check the colorimetric tolerance. If the color is within tolerance, exit with the calculated values. Otherwise, proceed to Step 3.
3. Calculate the DAC values needed to produce the color that appeared on the display, using the same computational method as in Step 1. Call these the "measured" values.
4. Subtract the measured values from the calculated values. Then, add the differences to the loaded values, yielding the "corrected" values. In other words:

   Corrected = Loaded + (Calculated − Measured)

5. Load the corrected values, measure the display, and check the tolerance. If the color is within tolerance, exit with the corrected values. Otherwise, return to Step 3.

Example: If the calculated DAC values are 100, 100, 100, but the color they produce on the display is out of tolerance and is what would have been expected for 110, 110, 110, the corrected values become (100 + 100 − 110 =) 90, 90, 90. If the color produced by 90, 90, 90 is also out of tolerance and is what would have been expected for 98, 98, 98, the corrected values become (90 + 100 − 98 =) 92, 92, 92. And so on.

It can be seen that the algorithm models the DAC-value error as a simple linear offset in three dimensions, i.e., as though the primaries' DC offsets ("black levels") have drifted. This is, of course, an oversimplification because of the typically nonlinear relationship between DAC value and display luminance. However, if the original solution is close to being correct, the linear approximation can be expected to at least reduce the error, which means the approximation will be even more adequate on the next iteration.

In practice, we have found that the algorithm works well (at least, for CRT monitors), even when the original solution errs considerably. For several years, we have been performing color vision experiments involving 210 colors (see Fig. 10) having luminances ranging from 18 to 36 cd/m^2. All 210 colors are adjusted each night to tolerances of $\pm 2.5\%$ in luminance and 0.0025 $u'v'$, using our algorithm plus PLVC. On the first night following calibration, we typically find that: (1) roughly 75% of the colors are out of tolerance; and (2) the program requires an average of slightly less than three measurements to bring each of these colors into tolerance. Thereafter, the program uses the previous night's corrected DAC values as the starting point and, again, typically requires an average of slightly less than three measurements to bring "bad" colors into tolerance. This is near ideal performance because all measure-and-adjust algorithms require a minimum of two measurements if the original color does not meet tolerance.

There are several additional points worth making about our algorithm. First, it is independent of the DAC-value computation method, as well as the method for assessing colorimetric tolerance. Second, it should work for any existing color display technology—not just CRTs. Third, damping should be included in Step 4 to guard against possible oscillation should the original DAC values be grossly wrong. (We have found that a good maximum permitted step size is 2^{N-4}, where N is the number of bits in each DAC.) Finally, to avoid infinite loops, two additional termination criteria are advisable:

1. If the corrected DAC values for a given iteration are identical to those that were measured in either the current or a previous iteration. (This indicates that the tolerances are too stringent for that color, given the display's gamut, the DG/display/spectroradiometer system's colorimetric resolution, and/or the system's temporal stability.)
2. If the number of measurements exceeds some (generous) user-specified limit. (This provides final protection against gross system instability.)

In both the above cases, the routine should exit with the measured DAC values that came closest to meeting tolerance.

Acknowledgment

We gratefully acknowledge the contributions of Gary L. Beckler, James H. Hill, and Donald H. Jones (Systems Research Laboratories, Inc.) to this project. Thanks are also due Robert J. Beaton (Virginia Polytechnic Institute and State University), who critiqued an earlier version of the manuscript.

1. William B. Cowan, *CIE calibration of video monitors*, (available from W. B. Cowan, National Research Council of Canada, Ottawa, Canada, K1A 0R6), 1986.
2. Willard W. Farley and James C. Gutmann, *Digital image processing systems and an approach to the display of colors of specified chrominance*, Technical Report HFL-80-2/ONR-80, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1980.
3. James C. Gutmann and Steven P. Rogers, *Displaying colors of specified chrominance on a color graphics display*, Technical Report TR 459-4, Anacapa Sciences, Inc., Santa Barbara, CA, 1982.
4. David F. Neri, *A simple calibration procedure for color CRT displays*, Report No. 1118, Naval Submarine Medical Research Laboratory, Groton, CT, 1988.
5. SAS Institute Inc., *SAS User's Guide: Statistics, Version 5 Edition*, Raleigh, NC, 1985, pp. 575-606.
6. David F. Neri, personal communication, 1988.
7. David L. Post and Christopher S. Calhoun, "An evaluation of methods for producing specific colors on CRTs," in *Proceedings of the Human Factors Society 31st Annual Meeting*, Human Factors Society, Santa Monica, CA, 1987, pp. 1276-1280.
8. Willard W. Farley, *Design and testing of a luminance and chrominance stabilization system for a computer-controlled color display*, Technical Report AAMRL-TR-87-027, Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, OH, 1987.
9. David L. Post, "U.S. Air Force Color Display Issues" (Technical Paper 861695), in *SAE 1986 Transactions*, Society of Automotive Engineers, Warrendale, PA, 1986, pp. 756-776.
10. William B. Cowan and Nelson Rowell, On the gun independence and phosphor constancy of colour video monitors, *Color Res. App. Supplement*, **11**, S33-S38 (1986).
11. David H. Brainard, Calibration of a computer controlled color monitor, *Color Res. App.*, **14**, 23-34 (1989).
12. William B. Cowan, An inexpensive scheme for calibration of a colour monitor in terms of CIE standard coordinates, *Computer Graphics*, **17**, 315–321 (1983).

Received December 8, 1988; accepted March 17, 1989.